

Weekly Report

2018.1029-2018.1104

1.This Week

Deep Learning Power Grid Project:

1.We apply symbolic representation SAX to process our time series data:

- we make trials to test the different bin width and alphabet length to choose the best k.
- we try to train the output data (after applying SAX) and it is not finished.

2.We explore the original dataset to check if the dataset itself is able to be separated by non deep learning methods:

- Talk to yuxin about the possible clustering methods. In theory, all clustering methods are applicable, but the problem is how to define the similarity between different samples (2000 * 2000 time series matrix) and he suggests to try DTW first. Another problem is that existing measures are all used for time series vectors, not matrices.
- We surveyed a bundle of time series similarity measures and analyzed which one is most suitable for our dataset (more than a dozen, I'm still organizing them into a word document).

Working Hour: (except nap and eat time)

8-9 hours / week day (except Monday)

6 hours on Saturday

Total Working Hour this week: 41 hours.

Other

- 1.Prepare the slice for the midterm report presentation.

Paper Reading

1.DQNViz: A Visual Analytics Approach to Understand Deep Q-Networks (VAST 2018)

This paper presents a visual analytics system and a visual design for sequence data to interpret the reward and action mechanism in reinforcement networks (the training process). The limitation is that this work only targets at Atari games, especially on the Breakout game and cannot be applied to more general reinforcement networks. One strong point of this paper is one of its insight that improves in controlling random actions in DQN models.

2.On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration



This paper proposes the flaws remained in applying recent time series mining techniques to real world problems because of data bias and implementation bias. Three main points are addressed: carefully choose the similarity measure in clustering and classification tasks, more than one dataset is needed to be tested and experiments should be designed to be free of the possibility of implementation bias.

3.A Symbolic Representation of Time Series, with Implications for Streaming Algorithms

4.An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback

These two works propose symbolic representations for streaming time series data. The aim is to reduce the dimensionality of large time series and reduce the computation cost on dealing with the output data. The first paper is more early than the second one. The second is a lower bounding approach and is stronger than the first one. Symbolic representations prove to be useful for tasks as clustering, classification, indexing, summarization and anomaly detection.

5.Hierarchical Classifier Design in High-Dimensional, Numerous Class Cases

This paper proposes a hierarchical classifier model (a hybrid decision tree classifier) for problems with only a small number of training samples but a large number of class labels. The question of feature extraction and the mathematical relationship between sample size, dimensionality, and risk value are also addresses in this paper.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	12.15	1.Finish training the model on the 2,000 node dataset. 2.Prepare to revise the method.
SQC Paper	-	1.Delayed

